

カリキュラム評価としての英語能力テスト

増原 仁美・木村 隆*

Proficiency Test as a Tool for Curriculum Evaluation

Hitomi MASUHARA and Takashi KIMURA

序

平成3年7月に大学設置基準の一部を改正する省令が告示・施行され、各大学は教育研究活動等の状況について、自己点検・評価を行うことに努めなければならないという義務が規定された。この規定は、大学が自らの教育研究水準の向上を図り、その大学の目的や社会的使命を達成するために、自主的・自律的に自らの点検・評価を遂行し、それによって打ち出された改善策を積極的に進めることが不可欠であることをわれわれ大学教育に関わる者に確認させようとしているのであろう。大学が自主的かつ自律的に点検・評価を行うべき項目の中には、当然のことながらカリキュラムも含まれている。カリキュラムというものは、本来常に評価を行うべきもののはずであるが、これを行うことは必ずしも容易ではない。今回の改正にカリキュラムの自己評価が含まれたのは、これが容易ならざるが故に今まで適切に行われてこなかったことの証左であろう。カリキュラム評価の方法論に関しては、欧米においてもまだ科学的研究の歴史は浅く、教育行政とも絡みながら徐々に整備・発展している状況であり、またわが国ではカリキュラム研究自体が大きく立ち遅れているのが現状である。このような状況に鑑み、本研究では、英語カリキュラムを評価するための一手段として英語能力テストを導入することの意義を考察しようと考えた。本稿では、英語能力テストとしてTOEFLを用いた事例を報告することによって、カリキュラム評価の意義や評価を行う上での問題点を指摘したい。

1. カリキュラム評価とは

カリキュラムという用語の定義は必ずしも一定していないが、安藤(1991)によると、「教育に関わる目的・目標、内容、方法・手順、時間配当・進度、教材・教具、および評価方法等を概括的に述べ教育の全般的方針を示したもの」とある。本稿においてはこの意味で論を進める。では、カリキュラム評価とはどのような概念を表すのであろうか。カリキュラム研究の領域を図式化したTaylor & Richards(1985)のモデル(図1)を利用することにより、カリキュラム評価がカリキュラム研究の中でどう位置づけられるかを以下に説明したい。

一連のサイクルの①の段階として、まず存在するのが教育理念やイデオロギーである。ここでは、教育の意味・目的・目標等を設定するために様々な価値判断がなされる。このようにし

*豊田工業大学

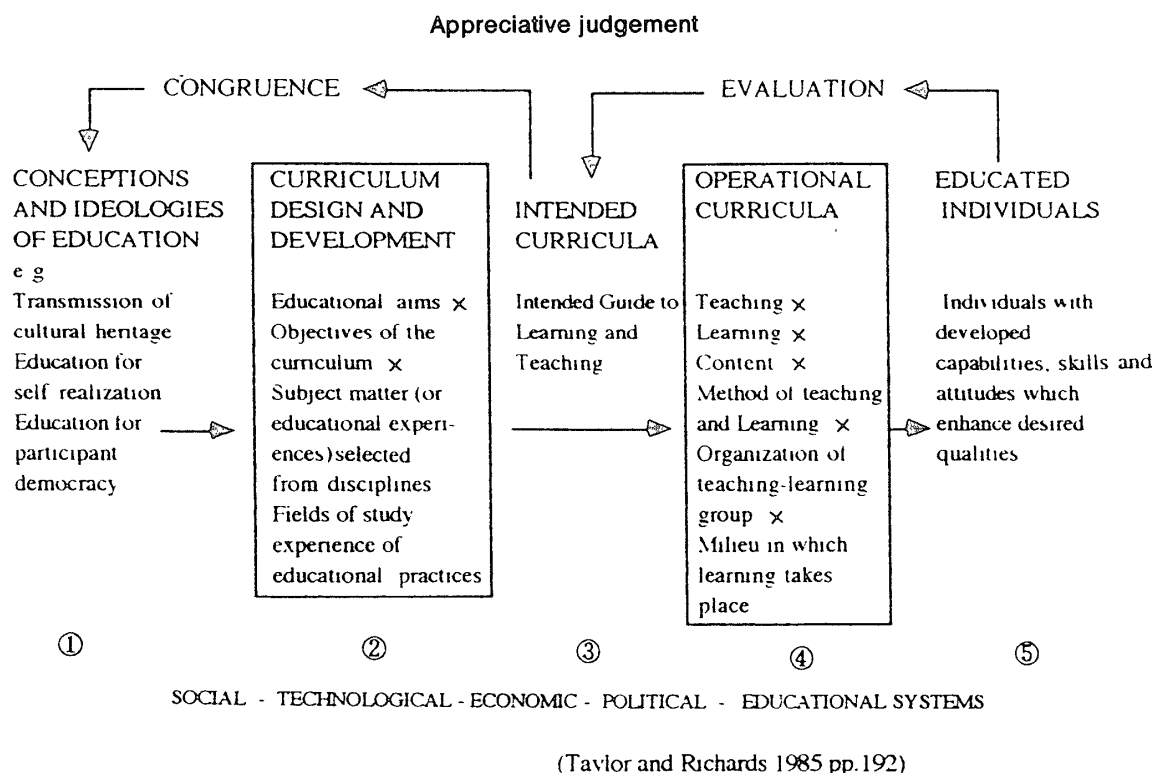


図1 The scope and concerns of curriculum studies: a tentative system model.

て行われた教育哲学的決定に基づいて、次に②の段階のカリキュラム開発や編成が行われるのである。カリキュラム開発は、既に実践されているカリキュラムの小規模な修正・改善から国や地域の教育機関等に於ける大がかりな新プロジェクトの開発まで含むのであるが、この段階では教育理念を具体化するために必要とされる様々な判断や開発方法の選択が行われ、当該教育機関で何のために何がどの様に教えらるべきかの提議がなされる。この提議に基づいて③の段階の、カリキュラム原案としての指導要綱が完成されるのである。無論、ここで翻って、①の教育理念やイデオロギーのレベルにおける価値判断がこの③のカリキュラム原案に具現されているかどうかの確認を必要とするのは言うまでもない。③から①に向かう矢印は、この確認のプロセスを示している。

しかし、カリキュラム原案が真に実体を持つのは、④の段階として実施用のカリキュラムが作られ、実行された時である。ここでは、カリキュラム原案が教授と学習の中での意味づけをされ、学習者に対する実践の中で提示される。即ち、原案は当該教育機関の現実的条件の中で練り直され、実施に関わる様々な決定を経て、具体的な形をとるのである。例えば、教授法の選択、教材の選択、シラバス作成、教授・学習環境の設定等、教育的判断を下さなくてはならない項目は非常に多い。そして、最後の段階である⑤には、一連のプロセスを経て完成されたカリキュラムの成果として、個々の学習者が存在するのである。この⑤の段階における学習者の反応や教育効果の測定・観察等を通して、再び③のカリキュラム原案に立ち戻り、意図された教育内容や達成水準が適切に実現されていたかどうか分析されなくてはならない。なお、②から④の段階にかけては、図に示されているように社会的・工学的・経済的・政治的・教育的システムが背景として考慮に入ってくる。

では、カリキュラム評価とは、この図式の中でどこに位置するのであろうか。この問いかけ

に対しては、視点によって2つの解答が考えられる。その一つは、⑤の個々の学習者から③のカリキュラム原案に至る矢印によって示された *evaluation* と呼ばれる部分である。またもうひとつは、このモデルの全体を評価する、*appreciative judgement* と呼ばれるプロセスである。この両方の視点は、カリキュラム研究の歴史的発展の中から生まれてきた3種のアプローチを理解することで説明される。第一のアプローチは、研究対象の教育活動とその結果を学習者個人がどう体得したかを測定することによって、行動目標の達成度を測ろうとするもので、結果の数値化や因果関係の追求を行うことを主眼にしたものである。Bloom et al (1956)や Chronbach (1963), Stufflebeam (1971)等が提唱者で、用いられた評価技法としては、行動目標の設定、テスト・アンケート・評定尺度の作成利用などがある。第二のアプローチというのは、教育を人間的かつ社会的活動と捉える立場で、自然科学的に数値で示された結果より学習者の体験の質や意味を問い、結果よりはむしろ過程の質を問題とした。(MacDonald 1973, Pratt 1980)。第三のアプローチは第一と第二アプローチの両方の要素を含むもので、教育活動の意図と実際の結果を観察や測定によって質的・量的に吟味しようとする立場である。カリキュラム評価を理論から実践に至るまで全面的に見直し、教育測定の専門家のほか社会学や政治学、歴史学者等との協力により分析・評価を試みることも提唱した(Stake 1976)。この吟味の過程は“*appreciation*”と呼ばれ、教育理念やイデオロギーが現実に行われている教育に具現されているかどうか等、教育過程全般の価値を問い、評価 (*appreciation*)することにより、修正や改善の必要性を決定 (*appreciative judgement*)しようとするものである。これら3種のアプローチは別個に存在するものというよりは、カリキュラムをより適切な方法で科学的に分析し得た場合には有機的に融合できると思われる。

評価という言葉は、現在までの先行研究に於いては、図1の *evaluation* というプロセスで示されている狭義の意味に使われることが多く、実施用のカリキュラムや結果としての教育効果が評価の対象とされたものが多い。本稿で事例として取り上げた実験も、英語能力テストを教授計画の前後に入れることによって教授効果の程度を分析するという点で、その範疇に入る。広義の意味でのカリキュラム評価である *appreciative judgement* に関しては、今後の発展研究の課題としたい。

2. カリキュラム評価のためのテスト

2. 1. テストが満たすべき要件

1で筆者らは、カリキュラム評価のアプローチのひとつが、学習者がカリキュラムの目標をどの程度達成できたかを測定することであると述べたが、そのような測定に用いるテストとはどのようなものなのであろうか。

まずはじめに問われるべきことは、そのテストの種類である。言語にかかわるテストはその目的や測定される言語項目などによって様々に分類されるが、カリキュラム評価の場合には、到達度テスト・能力テストのうちのどちらを用いるかが問題となる。

到達度テストについて J. B. ヒートン (1991) は次のように述べている。「校内試験のほとんどは到達度テストの形式を取っており、ある特定のシラバスの習得を示すことを意図した公開試験もまた、みな到達度テストである。これらのテストは生徒が学んだことになっているものに基づいているのであって、それは必ずしも彼らが実際に学んだことや、実際に教わったことではない。」また、同様に能力(熟達度)テストについては、「到達度テストが生徒が学んだことを振り返るのに対して、熟達度テストは前を見ており、生徒の言語能力を、彼らが将来行わ

されそうな特定の課題に照らして定義するものである。」と述べている。

テストを実施することによって、単に特定の科目のシラバスや教科書に基づく教授・学習活動の達成度についての情報を求めるだけならば、到達度テストを用いればよいが、カリキュラム評価では、あるカリキュラムによる教授・学習活動を通して、そのカリキュラムの目標そのものがどの程度達成されたかが分かり、そこから教授法や教材などのカリキュラム内容を評価・改善することができるようなテストでなければならない。たとえば、ある学習者が特定のシラバスや教科書に基づく到達度テストで高い得点を上げたとしても、それが即ち、学習者がカリキュラムの目標を十分達成したことになるとは限らない。というのも、そのシラバスや教科書自体がカリキュラムの目標と整合していない場合があるからである。このことから、カリキュラム評価に用いるテストは、その測るべきものが学習者の将来のニーズ分析に基づいて設定したカリキュラム目標と合致し、その目標に照らし合わせて、学習者が実際に身につけた学力を測定する能力テストであると言うことができる。

ところで、テストが一般に具備すべき性質として、垣田(1979)やハリス(1972)は、妥当性(validity)・信頼性(reliability)・採点容易性(scorability)・実施容易性(administrability)・経済性(economy)を挙げているが、カリキュラム評価のためのテストには、さらに(6)難易度の適切さ(7)測定結果の利便性(8)波及効果(backwash effects)の望ましさなども考慮されなくてはならない。

テストの妥当性とは、テストが測定しようとするものを正しく測定しているかどうかということである。カリキュラム評価のための英語学力テストならば、そのテスト問題の内容は英語カリキュラムの目標や内容を反映したものとなっていなければならない。また、信頼性とはテスト得点の安定度を意味する。カリキュラム評価ではカリキュラムの目標に沿った英語学力を継続的に測定する必要があるため、仮に同じ学習者が同じ日に何度受験しても、また一人の受験者の答案をいつ誰が採点しても同じ結果になるような、信頼性の高いテストが求められる。

採点容易性・実施容易性・経済性とは、ひとことで言うならばテストの実用性である。カリキュラム評価では、各教科の授業の中で担当教員が行う定期試験等と違って、はるかに多人数の受験者が関係してくることを前提にする必要がある。採点や実施の容易性はテストの信頼性にも大きな影響を与えるので、実施者の側の条件とも合わせて十分に考慮されなければならない。たとえば、行おうとしているテストは実施者の利用できる施設・設備で無理なく実施可能かどうか、監督者や採点者は何人位必要で、いつでも手配できるかどうか、テストの手順は明確に定められ、実施者が手際よく実施できるよう明解な指示が用意されているかどうか、などが問題となる。

テストの難易度が適切であるかどうかも大切なことである。カリキュラム評価では、英語学力の変化を時系列的に測定することが主になるので、その測定に用いるテストは、あるカリキュラムに基づいた授業を受ける前の学習者の学力のうちで最も低いものから、実施後の学習者の学力のうちで最も高いものまで、十分な信頼性を持って対応できるものでなければならない。もし受験者のうちの何人かでも、そのテストが測定できる最低の限界(floor)または最高の限界(ceiling)に達するようなことがあれば、それらの測定結果は受験者の学力を正確に反映していないことになり、受験者全体の測定結果に基づくカリキュラム評価を歪んだものにするからである。

測定結果の利便性というのは、テストによって得られた測定結果がカリキュラム評価にどの程度またどのように役立つのか、ということである。たとえば、実用英語検定試験いわゆる「英

検」や「ケンブリッジ英検」は、その結果が受験級の可否やバンド（評定尺度）という形で出てくるが、このようなテストは、その結果から学習者の小さな学力の変容を読み取ることが難しいのでカリキュラム評価には適さない。また英語カリキュラムの目標が Reading, Writing, Listening, Speaking の各技能領域全般に関わっているのならば、それらの領域のそれぞれについて測定結果が出るテストの方が望ましいことは言うまでもない。

テストの波及効果 (backwash effects) とは、テストが学習者の学び方や教師の教え方に与える影響力のことである。例えば、カリキュラム評価のテストをできるだけコミュニケーション能力重視のものにすることによって、日々の授業や学習の中で一層コミュニケーション活動を促すような刺激が与えられる。時田 (1991) は「テストからの悪い backwash effects には警戒しなければならないが、よいテストには、teaching の方向を変え、教師に自らの教え方を検討するよう促す有益な backwash effects があることも忘れてはならない」と述べている。

2. 2. 事例研究に用いたテスト

前節で、カリキュラム評価に用いるテストは、そのテストが測ろうとしているものがカリキュラムの目標と一致するような能力テストであると述べた。このことからすれば、カリキュラムの具体的目標が個々の大学によって異なるのが普通である以上、テストは個々の大学のカリキュラムに合わせてそれぞれ独自に作成されなければならないということになる。

しかしそうは言うものの、優れた英語能力テストの開発には長期間にわたる入念な準備や繰り返しの試行・データ分析が必要となり、筆者らの研究の範囲を超えてしまう。したがって今回の事例では、既存の英語能力テストを、カリキュラム評価のためのテストに求められる要件として前述した 8 項目に照らし合わせて吟味し、限られた条件の中で実施可能なものを選んで用いることにした。今回、授業方法の評価に用いることにしたのは、TOEFL (Test of English as a Foreign Language) の一種である ITP Level 2 TOEFL である。まずこのテストについて略述した後、採用に際して検討した事柄を述べたいと思う。

(1) TOEFL の目的と実施形態

T O E F L	International and Special Center Testing Program	アメリカ・カナダの大学への入学資格検定を目的として、世界各地の指定された会場で指定された日時に、TOEFL 本部が監督・実施する。単に TOEFL と言う時は、一般的にこのプログラムを指す。	
	Institutional Testing Program	ITP と呼ばれ、アメリカ・カナダへの留学希望者に対して英語教育を行う世界各地の語学学校などが、受講生のレベル分け等を目的として都合のよい日にそれぞれの施設やスタッフを使って実施する。難易度の異なる 2 つのレベルが用意されている。	
		Level 1	Level 2
		問題項目の難易度や項目数・実施時間など、テスト内容のすべてにおいて International and Special Center Testing Program と同一である。	過去に出題された問題項目のうち、難易度が比較的低いものを中心に選び出してテストを構成し、英語能力の十分でない受験者に対して信頼性の高い測定ができるようになっている。 また項目数が Level 1 に比べて少ないので、所要時間も短い。

TOEFL は、元来、英語を母語としない人々を対象として、アメリカ・カナダの大学に入学

しても授業について行けるだけの総合的な英語能力を身につけているかどうかを判断するために開発されたテストであるが、現在では前ページの表のように実施形態やレベルが多様化し、語学学校などにおけるレベル分けにも用いられるようになってきている。

今回の事例で Level 2 を用いることにしたのは、受験対象者の英語能力レベルが全体として Level 2 の難易度に最も適合すると予想されたこと、正規の授業時間 (90分) 1 コマの中でテストを実施する必要があったこと、できるだけ経費を抑える必要があったことなどの理由による。

(2) テストの構成

ITP Level 2 TOEFL を含めてどの TOEFL も、Section 1: Listening Comprehension, Section 2: Structure and Written Expression, Section 3: Reading Comprehension and Vocabulary という 3 つのセクションで構成されている。各セクションとも 4 肢選択のマークシート方式で解答し、コンピュータを使って採点される。セクション 1 では、自然な速度で話されるアメリカ英語が録音テープで提示され、口語英語でよく使われる語彙や構文の理解度の他、英語を母語としない者にとっては理解が難しい発音・イントネーションを聞き分ける能力もテストする。セクション 2 は構文と作文力に関するセクションで、標準アメリカ英語の重要な語法や文法についてその運用力を測る。出題形式は、文完成問題と誤文訂正問題である。セクション 3 は、受験者がアメリカの大学の授業について行けるだけの読解力と語彙力を身につけているかどうかを測定する。問題文の題材は一般的な教養に類するもので、特別な分野の研究者や特定の国家に属する者に特に有利にならないように配慮されている。ITP Level 2 TOEFL の問題項目数と解答時間は、セクションごとの解答方法の指示を含めて、セクション 1 が 30 問 (22 分)、セクション 2 が 25 問 (17 分)、セクション 3 が 40 問 (31 分) である。

(3) スコアについて

TOEFL の総合得点は、中心点を 500 点として上下 300 点の幅をとり、200 点～800 点の間で表示されるようになっている。また各セクション別の得点は 20 点～80 点の間で示される。ただし今回の事例で使用した ITP Level 2 TOEFL は、難易度が抑えられているため、総合得点の範囲が 200 点～500 点、セクション別得点の範囲が 20 点～50 点となっている。TOEFL の開発・実施機関である Educational Testing Service (ETS) ではテストフォームごとの難易度の違いから生じるスコアの変動を補正するため、毎回のテスト結果に変換スコア算定式をあてはめ、セクションごとの正答数を変換スコアに変えて示す方法を取っている。

さて、この TOEFL をカリキュラム評価に用いようとするれば、まずカリキュラム評価テストとしての妥当性が問われよう。TOEFL が留学志願者の英語試験としていくらか妥当性の高いものであっても、カリキュラム評価のためのテストとして妥当なものかどうかは、改めて検討されなければならない。カリキュラム評価に用いるべきテストは、そのテストが測ろうとしているものがカリキュラムの目標と一致するようなテストであることは繰り返し述べてきたが、そうすると大学英語カリキュラムの目標とは何かが問題となる。個々の大学の英語カリキュラムが具体的に目指すものは様々だろうが、わが国大学英語教育が目指すべき方向として、大学英語教育学会 (JACET) (1992) は次のような教育改善のための基準項目を提案している。「大学における外国語教育の目的は、当該言語の言語運用能力の養成をはかることによって異言語文化を体験し、異なる人間の世界を発見し、人格的な陶冶をはかることにある。従って、大学における外国語 (英語) 教育の実施にあたっては、言語運用能力つまり人間の言語によるコミュニケーション能力の養成を第一とすべきである。」また上掲書には、学部教育における具体的教育目標の設定に関して、「外国語によるコミュニケーション能力の養成に関しては、a. 語るべ

き内容に関する教養, b. 討議の形式や対人関係等への適応性, c. 話題の専門領域に関する知識構造と語彙への習熟, 等と関連した目標を具体的に設定すべきである。」と述べられている。これらの基準項目からは、今後の大学英語教育が伝達内容の受容能力だけでなく発信能力も目指していくべきことが読み取れる。一方 TOEFL は、留学志願者に対してアメリカ・カナダでの大学授業を受けるための言語レディネスを測るという本来の目的から、どちらかと言えば受容面の言語技能に偏っている。ETS は、TOEFL と共に実施して受容面の測定を補足するためのテストとして、Test of Written English および Test of Spoken English を既に開発・実施しているが、そのどちらも ITP には含まれていない。また、TOEFL に用いられる英文テキストの種類や話題はアメリカ・カナダの大学生活に限定されたものが多い。たとえばセクション 1 では、体育の授業で泳力によってグループ分けをするような場面や、学生寮での避難訓練の場面が取り上げられていたり、セクション 3 では、大学教科書で典型的に見られるような説明的(expository)な文だけが使われていたりする。

これらの点を考えると、TOEFL はわが国大学英語カリキュラムの目標の一部について対応しているに過ぎないので、一般的にはカリキュラム評価のためのテストとして妥当性が不十分と言わざるを得ない。ただ、個々のカリキュラムの具体的な目標によっては、より妥当になることもあるし、カリキュラムのうちの限られた一部を評価するような場合には、適切と見なし得ることもあると思われる。

さて、妥当性の点でカリキュラム評価には最善のテストではないにもかかわらず、今回の事例で取って TOEFL を採用することにしたのは、Level 2 TOEFL が前述のテスト要件のうちの、信頼性、採点容易性、測定結果の利便性において優れていると考えたからである。

TOEFL の信頼性は大変高いが、それは ITP Level 2 TOEFL にも当てはまる。Educational Testing Service(1991)によれば、セクション 1・2・3 の信頼性はそれぞれ 0.76, 0.80, 0.85 であり、総合得点の信頼性は 0.91 である。採点については、解答のすんだマークシートを ETS の日本代理機関に送付するだけでよいので全く手間はかからない。2 週間後には、受験者各人に手渡す個人成績カードと受験者の成績一覧表が返送されてくる。測定結果の利便性に関しては、総合得点の他に技能領域別得点も出るので、各技能領域についてカリキュラムのバランスを評価するのに役立つ。また測定結果は、各セクションでは 20~50、総合得点では 200~500 の数字で表されるので、受験者の英語能力のわずかな変動を容易に把握することができる。また、見逃してはならないのが TOEIC との得点換算である。TOEIC (Test of English for International Communication) は英語を母語としない人を対象として、国際語としての英語コミュニケーション能力を測定するために ETS が開発したテストだが、ETS からは TOEIC 得点と TOEFL 得点との間の換算式が発表されている。わが国では多くの企業が TOEIC を社員の英語能力の評価などに利用しており、またわが国の TOEIC 実施代理機関からは、団体受験をした企業の大卒新入社員の平均得点や、企業が大卒新入社員や部門別海外要員に期待する得点などについての調査結果も発表されている。

3. 事例研究

3. 1. 目的と方法

本学英語科の現カリキュラムにおいて、リーディングの指導はそれぞれ通年 2 単位の英文講読Ⅰ・Ⅱ・Ⅲ・Ⅳで行うことになっている。英文講読Ⅰ・Ⅱは一年時に、英文講読Ⅲ・Ⅳは二年時に履修し、内容の区別としては、Ⅰ・Ⅲで「精読」(インテンシブ)を、Ⅱ・Ⅳでは「多読」

(エクステンシブ) を教えるという基本デザインになっているのである。設置基準の改訂に伴い、カリキュラムの見直しが検討されている現在、英語科のカリキュラムを改善・修正或は新しく開発していくための判断材料として、教授活動の教育効果を分析することの意義は大きいと思われる。従って、筆者らは以下のような設定で事例研究を行った。

研究目的： 英文読解を行う時に、先行研究で有効とされているストラテジー（読み手が英文を効率よく読解するために使う手順や方略）を教授すると、英語読解力（本実験では TOEFL のリーディングセクションの得点）が向上するかどうかを調べる。

研究仮説： ストラテジー訓練を施した実験群は、同じ条件下でストラテジー訓練を含まない統制群に比べ、より読解力が向上する。

被験者： 実験群 — 44名 統制群 — 49名（両群とも短期大学部英語科1年）

期間： 1992年5月～7月（90分1コマ/週）

内容： 実験群 — ストラテジー訓練
統制群 — ストラテジー訓練を含まない多読訓練

教材： ストラテジー訓練 — 訓練シラバスに基づく教材
多読訓練 — Parker, D. H. SRA Reading Laboratory 2c,
Macmillan / McGraw-Hill

実験群に対して実験期間中に教授したストラテジーは、未知語の推測・スキミング・パラグラフパターンの識別・代名詞及び言い換え表現の4種類である。実験群の教授法は、学習者にストラテジーの存在を気づかせ、より有効なストラテジーを説明し、ストラテジーを習得するように構成された練習問題に取り組ませるといったものであった。

一方、統制群に使用したSRA Reading Laboratoryは、個々の学習者が自らの理解能力・速度・興味に合わせて系統的に多種類のレベル分けされた教材を読み進むように構成された多読教材集である。教授者の役割は、個々の学習者に教材の意図や使い方を説明するなどの補助をするのみで、知識の伝達は行わない。なお、SRAの教材には本文の他に内容の理解度を測る問題と未知語の推測等のストラテジー強化のための練習問題がついているが、統制群には内容の理解に関する問題にのみ取り組むよう指示を与え、ストラテジー訓練は一切行っていない。

3. 2. 結果

1. 等質性の検定

図2 プリテスト得点の度数分布（実験群）

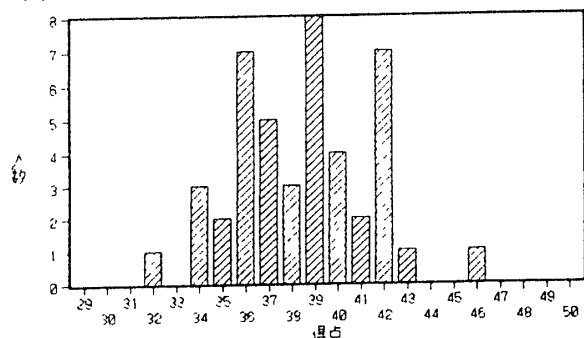


図3 プリテスト得点の度数分布（統制群）

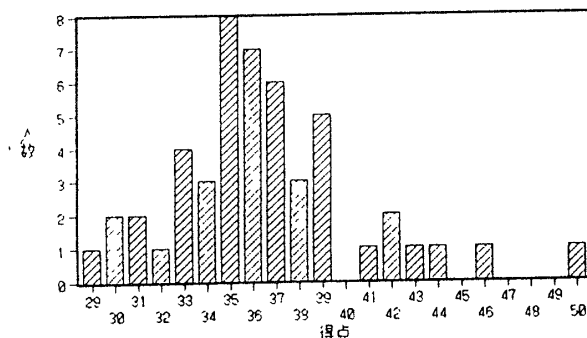


表1 z 検定の結果

		平均	標準偏差	z 値	F 比
実験群	(N=44)	38.46	2.896	2.542(NS)	2.036(p<0.05, df=43.48)
統制群	(N=49)	36.59	4.137		

2. プリテストーポストテスト間の得点変化

図4. プリテストーポストテスト間の得点変化

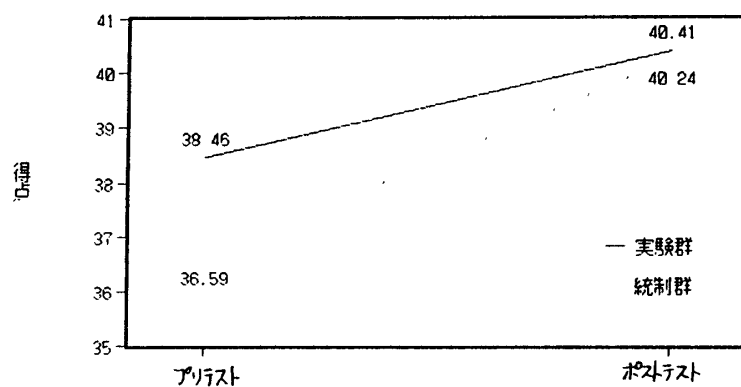


表2 群内の得点変化

	プリテスト		ポストテスト		伸び	t 値
	平均	標準偏差	平均	標準偏差		
実験群 (N=44)	38.46	2.896	40.41	3.473	1.955	4.47(p<0.001)
統制群 (N=49)	36.59	4.137	40.24	4.439	3.652	6.48(p<0.001)

3. 実験以外の要因の確認

表3 被験者が実験期間中に大学の授業以外で行った英語学習

	実験群	統制群
*語学学校・カルチャーセンターなどに通って勉強した	6 名	6 名
*テレビやラジオの英語講座を視聴した	9 名	13 名
*通信教育を受講した	1 名	0 名
*英字新聞を週1回以上読んだ	1 名	4 名
*英文書籍を1冊以上読んだ	2 名	10 名
*その他の方法で勉強した	3 名	4 名

比較実験の厳密な手続きとしては、事前事後の変化の分析に入る前に、実験群と統制群が等質であることを確認しなければならない。図2・図3は、それぞれ実験群と統制群の事前テスト結果の度数分布をグラフで表したものである。図の2つの群の分布の違いに表れている通り、平均点と分散の差を統計的に検定した結果（表1参）、両群は等質でないという結果がでた。従って、以下ではそれぞれの群内の伸びを話題の中心にし、群間の比較はあくまで参考程度にとどめることになるが、現実の教育現場ではクラスの枠や教育的配慮から実験の厳密性のみを優先するべきでないためこのような結果になったのもやむを得ない。

図4のグラフが示しているのは、それぞれの群の訓練実施前後の平均点の向上の様子を表にしたものである。どちらも群内の伸びは表2にある通り統計的検定に於いて有意であった。研究仮説では、ストラテジー訓練を施した実験群の方が向上の度合いが大きいであろうと予測したが、結果は意外にも多読訓練を行った統制群の伸びの方が明らかに大きかった。

両群の伸びの差が本研究の教授法の違いのみによるものであるということを立証するために、学習者が研究期間中に大学で履修していた英語関連科目と、大学の授業以外に受けていた英語学習を調査した。その結果、学内では10教科に及ぶ英語関連科目中、一般教養科目としての「英語I」一科目で教授者とテキストが異なっていた以外は両群とも全く同じ教育を行っていた。また、学外での学習に関しては、海外語学研修を行った者は皆無であり、その他の結果は表3の通りであった。

3. 3. 考察

本実験の実証研究そのものとしての厳密な考察は別稿（増原他執筆中）に譲るとして、カリキュラム評価の観点からのみ、結果が示唆するものを取り上げると以下の様になる。

まず、実験群に与えたストラテジー訓練の妥当性であるが、向上が見られたというものの、群内の伸びを比較すると多読訓練を受けた統制群の伸びに比べ向上の度合いが小さい。これはストラテジー訓練の意義を否定するものなのであろうか。母語に於ける多くの研究は元より、外国語に於けるストラテジー教授効果の先行研究（Barnett 1988, Kern 1989, Carrell et al 1989）等を見ても、ストラテジー訓練の効果は近年ますます認められてきている。将来、類似の研究デザインで、教材・教授法・使用テストの選択・教授期間等を変化・改善させながら、標準化された英語能力テストを使って学習者の向上の度合いを測るという研究を継続していく中で、より明確な解答が得られるはずである。

一方、多読訓練を行った統制群の伸びはどのように解釈されるのであろうか。両群の教授法を学習者の観点から比較してみると、最も異なっていたのは英語のインプットの量であったと思われる。90分一コマの講義中、統制群の学習者は中間レベルの教材でも750～850語程度の英文からなるテキストを自分のペースで読む。中には一時限内に2つの教材を読みこなす意欲的な学習者も少なくなかった。それに比べ、ストラテジー訓練を受けた実験群の学生は、練習問題で多くの英文に触れるとはいえ読解の過程を内省したり未知のストラテジーを意識的に学習し使用する分、読解速度や触れる英文の量が少なからざるを得なかった。つまり、高校を卒業したばかりで訳読方式の読解しか知らない学習者にとって「読解の目的やテキストの種類・構成等を分析的に見ながら自覚的に英語を英語のまま処理する」というストラテジー使用に支えられた読み方は、通常の英文読解の作業に加え、自らの読解過程を意識することが要求されるので負担が大きすぎたのかも知れない。この考察は、ストラテジー訓練の有用性を認めるものの、読解力がついたという実感が薄いという事後アンケート調査の結果にも裏付けされる。

統制群の向上の軌跡は、本実験の学習者の現発達段階に於いて英文を英文のまま読みこなす経験をさせることの重要性を示唆しているといえないだろうか。言い換えれば、本学習者がストラテジー訓練を受け入れるだけのレディネスがない場合は、大学の初期の読解カリキュラムには、学習者の意識を高めるストラテジー訓練よりは、むしろ英文に慣れさせ英文に対する親近感を持たせる様な多読訓練が適しているのかも知れないのである。

4. 結 論

4. 1. カリキュラム評価に於けるテストの発展的使用—その可能性

3において筆者らは、カリキュラム評価の一環として英語能力テストが教授効果の測定にどのように使われ、その分析が教授法やカリキュラム改善にどう利用されうるかという問題を、事例研究を考察する中で考えてみた。事例のような教授効果の測定の他にも、テストの客観的データの分析はカリキュラム評価の様々な場面で様々な関係者に有益な示唆を与えられると思われる。以下では、英語能力テストがカリキュラム評価にどういう場面でのどのように利用されうるのか、また、その結果のデータが英語科・教授者・学習者にもたらす示唆にはどのようなものが含まれるのかを、テストの発展的使用の可能性をいくつか挙げる中で考えてみたい。

1) 各年度入学者の学力の記録

年度ごとの入学者の学力変化を比較し、大局的なカリキュラム修正・改善の資料とすることができる。例えば、入学者の各スキル別発達段階を考慮して必修・選択科目を構成し直したり、導入時の教育内容・教授法・難易度を考慮する等多くの利用法が可能と思われる。また、上記事例では実験群と統制群が等質でなかったが、クラス分けの参考資料としての利用価値も多いにあると思われる。教育の目的に合わせ、適切な内容や順序で教育が始められるということは、教授者にとっても学習者にとっても好ましい。また、長期にわたって各年度ごとの入学者の学力推移を記録する事で、入学者の学力の傾向もつかめる。

2) カリキュラム全体における教育効果の測定

上記の事例では1科目を教授するために用いた2つの教授法の効果を研究したが、最終的にカリキュラム全体が学習者の能力を向上させているかどうかを定期的に評価する必要もあるであろう。例えば、英文講読であれば英文講読Ⅰ・Ⅱ・Ⅲ・Ⅳそれぞれ別個の効果を事例のように調べるだけでなく、講読のコース全体で一定の期間後どのようなリーディングの力が向上したかを見たり、リスニングや作文も同様に定期的な評価を行うことで、カリキュラムの有効性が問われなくてはならないであろう。

また、語学教授は集中的に行う方が効果的との考察に基づいて、多くの教育機関では90分間の授業が1週間に1度行われる従来の実施形態を、集中訓練型に編成し直す傾向が見られるようになった。実施前にミニプロジェクトとして「集中型での教授」と「従来型」の教育の効果を比較する等、大幅な変更を要するためにすぐには実施が難しい場合のテストケースの評価に利用することも可能と思われる。

3) 習熟度別クラス編成資料

学習者の発達段階によって生じる学習差を考慮すると、語学訓練の目的によっては習熟度によるグループ分けが必要となる。テストのデータは学習開始前や、しばらく教育を行った後の再調整等の有効な判断材料になると思われる。また、社会人入学者・帰国子女・編入生等の学力も、統一した基準で位置づけることが可能になる。

4) 単位の弾力的取得のための基準資料

新設置基準では、国内外での教育機関による単位を短大で15単位まで認めるとなっているが、その際の免除教科等の選択資料にもなり得るのではなかろうか。例えば、英米からの帰国子女の場合は、テストにおいてリスニングの得点が高ければLLの単位を認め、むしろ海外で学習できなかった教科を履修するよう指導する等の対応も考えられよう。

5) 国際化に対応する英語力評価尺度の採用

国際化が叫ばれ、英語科でも毎年長期・短期の留学生を送り出している。また、企業では海外との通信・交流がますます不可欠になっている昨今、TOEICを入社時や入社後の研修の一部として受験させる企業が増加しつつあることも2. 2.において述べた。このような情勢において、国際的に高い評価を受けているテストを採用する利点は多いと思われる。例えば、本大学での英語力評価の結果を英語圏の姉妹校への編入資料としたり、カリキュラム評価の国際的比較なども将来の研究交流として可能になるからである。

6) 学習者へのフィードバック

教育の究極的目的の1つとして、学習に自覚的・自主的に取り組む主体としての学習者育成が含まれると思われるが、この観点から考えると、能力テストの結果は、教育機関・教授者がカリキュラム評価の手段として使うだけでは不十分であろう。即ち、教育機関は、学習の主体である学習者自身が自らの発達段階を理解し、学習によって得られる向上を科学的に把握することによって自らの学習をより望ましい方向へ制御できるような資料を提供していく義務があるのではなかろうか。上記事例においても、学習者にはTOEFL受験の意義や目的、またスコアの読み方やスコアの利用法等を説明した印刷物に加え、TOEFLに表れた学習の変化を各自がグラフに書き込むための用紙が配布された。定期的にこのような評価を行い、学習者に自らの学力の変化を知らせると同時に、必要に応じて個々の語学力カウンセリングの実施を図ったり、将来は更に進めて、学力向上のための個別LLの使用・リーディングラボ・矯正のためのスピーチクリニック等の個別指導も、カリキュラムの中に含むべきと考えられる。

4. 2. 総合的なカリキュラム評価をめざして

1において、カリキュラム評価として狭義の evaluation と広義の appreciative judgement の2種類を示した。本稿では、事例とこれからの展望を述べる中で、主として量的な evaluation の在り方を考えてきたわけだが、最終的にカリキュラム評価を正しく行うためには、カリキュラム全体に関しての appreciative judgement が不可欠であることに気づく。なぜなら、evaluation が妥当性を持つためには、その前提として、1で述べた Taylor and Richards の図の①～⑤の一連のプロセスにおいて、それぞれの段階の間の質的整合性が保証されなくてはならないからである。即ち、evaluation は全体の appreciative judgement の一部として機能して初めて意味を持つのである。では、appreciative judgement はどのように行うべきなのであろうか。評価の主体は誰なのであろうか。新しいカリキュラムの形態はどのようなものであるべきなのであろうか。今後のカリキュラム評価(appreciative judgement)の発展研究が多いに待たれる。

参考文献

- 安藤昭一, 1991, 「英語教育現代キーワード事典」, 増進堂, p. 57.
 青木宗也, 1992, 「大学・短大の自己点検, 自己評価」, エイデル研究所.
 Barnett, M. 1988. "Reading through Context : How Real and Perceived Strategy Use Affects L₂ Com-

- prehension". *The Modern Language Journal*, 72, ii
- Bloom, B. S. et al, 1956. *Taxonomy of Educational Objectives: Handbook 1: Cognitive Domain*. London: Longmans.
- Carrell, P. L., Pharis, B. G., & Liberto, J. C. 1989. "Metacognitive Strategy Training for ESL Reading". *TESOL Quarterly*, vol. 23, no. 4, 1989, 647-678
- Chronbach, L. J. 1963. "Course Improvement through Evaluation". *Teachers College Record*, 64, 672-683.
- 大学英語教育学会 (JACET). 「大学設置基準改正に伴う外国語 (英語) 教育改善のための手引き (1) . 大学英語教育学会. 1992. pp. 6-7.
- Educational Testing Service. 1991. *TOEFL ITP*. Educational Testing Service.
- デイビッド・ハリス著・大友賢二訳, 1972, 「英語の測定と評価」, 英語教育協議会 (ELEC).
- J. B. ヒートン著・語学教育研究所テスト研究グループ訳, 1992, 「コミュニケーション・テストング」, 研究社出版, p. 254.
- Hughes, Arthur. 1988. "Achievement and Proficiency: The Missing Link?" In Hughes, A (Ed.) *Testing English for University Study*. Modern English Publications in association with The British Council. p. 40.
- Hughes, Arthur. 1989. *Testing for Language Teachers*. Cambridge University Press.
- 梶田穀一, 1979, 「学校経営のための教育評価」, 永野重史・東 洋編著『教授・学習・評価』, 教育学講座第5巻, 学習研究社, p. 157.
- 垣田直巳編, 1979, 「英語教育学研究ハンドブック」, 大修館書店, p. 499.
- Kern, R. 1989. "Second Language Reading Strategy Instruction: Its Effects on Comprehension and Word Inference Ability", *Modern Language Journal*, 73, 135-149.
- Lemke, E. & Wiersma, W. 1976. *Principles of Psychological Measurement*. McNally College Publishing Company.
- MacDonald, B. 1973. "Humanities Curriculum Project." *Evaluation in Curriculum Development: Twelve Case Studies*. London: Macmillan
- 日本私立大学協会, 大学設置基準等の一部を改正する省令新旧対照条文. 平成3年6月
- Pratt, D. 1980. *Curriculum Design and Development*. New York: Harcourt Brace Jovanovich
- Stake, R. "The Countenance of Educational Evaluation". *Teachers College Record*, 68, 7, 523-540
- Stufflebeam, D. 1971. *Educational Evaluation and Decision-making*. Irasca: Peacock
- Taylor, P. H. & Richards, C. M. 1985. *An Introduction to Curriculum Studies*. Nfer-Nelson
- 時田育子, 1991, 安藤昭一編「英語教育現代キーワード事典」, 増進堂, p. 257.
- 安彦忠彦, 1985, 「教育評価研究とカリキュラム」, 安彦忠彦編『カリキュラム研究入門』勁草書房, p. 183.